Analyzing Science Operations for the Search for Life as part of a Multi-Year Robotic Campaign to Exp

First Year Report

Sponsor: Applied Information Systems Research Program,
National Aeronautics and Space Administration
Project Number NNG05GA51G

Geb Thomas, PI
Nathalie Cabrol, NASA Ames
David Wettergreen, CMU
Robert Anderson, JPL
Peter Coppin, CMU
Edmond Grin, Ames

Contact Information
Geb Thomas
Department of Mechanical and Industrial Engineering
The University of Iowa
Iowa City, IA 52246
Phone: (319) 335-5936
Fax: (319) 335-5669
URL: http://grok.ecn.uiowa.edu

February 27, 2006

Summary

In the austral spring (September-October) of both 2004 and 2005, a multi-institutional, multi-national effort led by Carnegie Mellon University conducted an extended duration robotic exploration of the Atacama Desert in Chile to develop technologies and methods for upcoming NASA missions that will search for evidence of past life on Mars. Our project leverages this large-scale effort to measure and improve the effectiveness of robotic science operations. This research will extend current analysis of rover-mediated geology to rover-mediated habitat characterization. The work emphasizes the effects of different data collection and display techniques on the science team's conclusions. The principal hypothesis of this research is that the quality and reliability of scientific conclusions regarding past or present life in arid environments is dependent on the type of evidence collected by the rover, the scientists' data analysis techniques, the processes used by the scientists' to form and share hypotheses and conclusions, and the science operations software. The principal objectives of this research, each specifically associated with AISR program objectives, are to: 1) reduce mission development time by analyzing how scientists characterize a habitat, 2) reduce mission development risk by identifying mission-critical and problematic analysis tasks, 3) increase science return from the data by analyzing long-traverse science collection strategies, and 4) increase data return by refining the science interface to improve analysis effectiveness.

This research will analyze the processes used by astrobiologists and geologists when searching for signs of life in a Mars-like environment. Our previous and ongoing work with robotic geology has successfully characterized limitations in scientific interpretations caused by rover sensors, differences in scientists' interpretations, and limitations of the science interface. These limitations were identified and studied using perceptual experiments in which scientists analyzed sample images and physical specimens. In addition, transcripts of scientists participating in a simulated rover field experiments have been examined to further understand these limitations. This project will quantify analyst and instrument limitations that could affect the success of future missions in the search for life on Mars and will develop mitigating strategies to avoid inappropriate conclusions regarding the presence of life on Mars.

Introduction

The science information interface that displays rover-collected data during a planetary exploration mission is the science team's principal window to another planet and is therefore critical to the mission's success. Previous rover missions and field tests have demonstrated that the design of both the rover's instruments and the science information interface affect the quality of the scientific interpretation of the remote environment. Our recent research has isolated and quantified specific differences in scientific interpretation between information presented in a picture and information directly perceived. What is not yet established, however, is the degree to which limitations in specific science analysis tasks can affect the scientific success of a rover mission and whether strategies to mitigate these limitations can significantly improve scientific success. This is a critical research issue because if problems associated with rover-mediated scientific interpretation are not identified and corrected, they could mislead time-stressed mission scientists toward overconfident, under confident, or erroneous scientific conclusions. These are serious problems that threaten to jeopardize NASA's success in robotic planetary exploration.

The long-term goal of this research is to improve the effectiveness of rover operations by identifying the operational limits of scientific interpretation and discover means to eliminate these limitations. The objective of this research project, which is the next step towards attaining our long-term goal, is the analysis of the tasks conducted by a science team during a rover mission and the refinement of their science information interface to enhance science operations. Specifically, we are analyzing the operations of a science team engaged in a field campaign in the Atacama Desert, characterizing which analysis tasks are the most limiting to the success of this campaign, designing and implementing solutions that eliminate these limitations, and confirming that the solution may be generalized to other remote exploration experiences. We need to carefully and exhaustively characterize the analysis tasks conducted by the science team in order to isolate the connection between any misleading or erroneous scientific conclusions and the details of the analysis of the original rover data that led to the misconception. Once the connection between the analysis and the misleading conclusion has been made, techniques may be developed that reduce the chance that similar mistakes will be repeated in the future. This research will lead to rover science operation software that mitigates common and egregious interpretation errors and improves the efficiency of time-consuming tasks. Our research team is particularly qualified to conduct this research because of our extensive experience building rover science interfaces, building exploration rovers, running rover field experiments, participating in planetary missions, and exploring remote and planetary terrains.

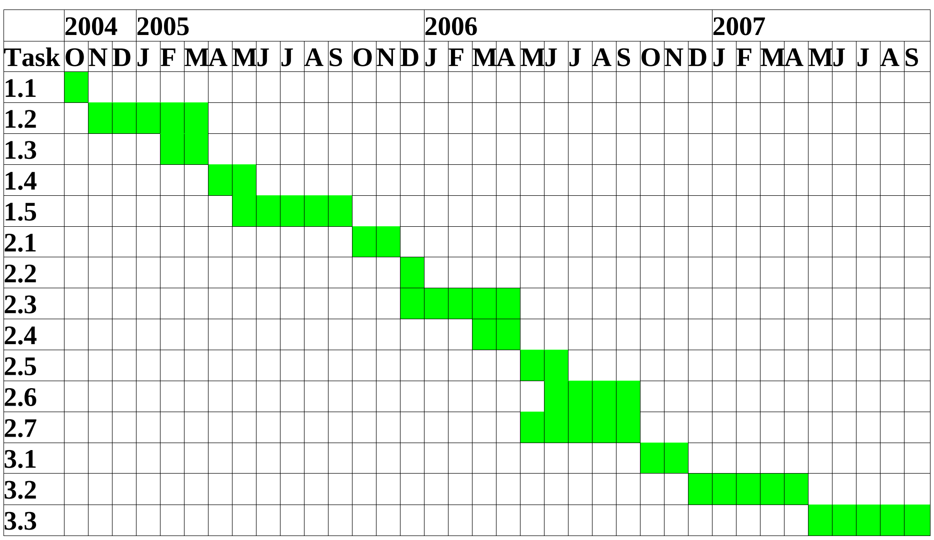
2.1 Objectives and Expected Significance

1. **Reduce mission development time by analyzing how scientists characterize a habitat.** Determine what tasks scientists conduct during a search-for-life rover mission, what information is used in these tasks, the time devoted to each analysis activity, and the reliability of the conclusions that may be drawn from these tasks. The expected significance of this objective is that future missions may be developed more quickly based on clear definitions of what analysis tasks will be conducted, what data each analysis requires, how long each analysis takes, and the limitations of the conclusions that may be drawn from each analysis.
2. **Reduce mission development risk by identifying mission-critical and problematic analysis tasks.** Evaluate and compare the observations made by control-room scientists during a search-for-life rover mission with observations of the same area made by scientists in the field and laboratory. Determine which control room conclusions were accurate with appropriate confidence levels and which conclusions were false or made with inappropriate confidence levels. The expected significance of this objective is the identification of analysis tasks based on rover-collected data that could mislead scientists' conclusions regarding evidence of life on Mars.
3. **Increase science return from the data by analyzing long-traverse science collection strategies.** Assist in determining the effect of alternative rover data collection

strategies during a long (1 km +) autonomous rover traverse, comparing the effect of collecting data samples at regular intervals, collecting data samples at locations specified by scientists in advance, and collecting data samples at locations determined by the rover's autonomous subsystems. The expected significance of this objective is the validation that the proposed analysis techniques can help to optimize rover exploration strategies to ensure the greatest scientific return for each rover mission.
4. **Increase data return by developing science information interface strategies that improve analysis effectiveness.** Refine an existing science information interface to specifically support scientists in making accurate and efficient observations and conclusions. The expected significance of this objective is the development of improved science operations software and the quantification of science performance gains resulting from the analysis and development supported in this proposal.

2.6.3 Timeline

The timeline of the original proposed project is outlined below. According to this timeline all of the tasks up to task 2.2 should be complete by the date of this report. Nearly all of these tasks have been completed. The exception is that two analysis tasks from the 2004 LITA expedition are still ongoing because their results must be compared with the results from the 2005 LITA expedition to determine their significance. There are also several rock and soil samples from the 2004 expedition that are currently undergoing commercial laboratory analysis.



2004-2005
1.1 Participate in the Atacama field campaign, two weeks of operations in Pittsburgh, collect audio and video tape recordings, as well as computer logging data indicating which data was used by whom at what time. Send daily reports of science observations and conclusions to the field for confirmation by the instrument scientists, including sample collections. (All) This task was successfully completed. Each day of operations 2 members of the Iowa team were present in the control room throughout the test. Each scientist wore a lapel microphone and six cameras were mounted in the ceiling to record

the position of all the scientists. Each day of the mission we interrupted the scientists to ask them to make 3 to 5 specific, testable hypotheses. These hypotheses were then tested in the field either immediately by team members in the field, or afterwards when members of our team visited the Atacama.

1.2 Create transcripts of the audio recordings from science team leaders. Use video recordings and data logs to trace which teams were working together at each point in the mission and to which data they were referring. Analyze the person-hours that accompanied each individual analysis task in order to find analysis bottlenecks among the scientists. (Thomas) This task was successfully completed. Full transcripts of the science team activities were collected and reported in the literature (e.g., Pudenz, Glasgow, Thomas, et al., “Searching for a Quantitative Proxy for Rover Science Effectiveness.”). However, this data set has proven to be very rich and analysis is still continuing.

1.3 Review transcript analysis, task analysis and interpretation network for domain-specific accuracy. Resolve controversial observations from laboratory studies of samples from the field. Trace any analysis errors revealed in the field back through the transcript to find the origin of each interpretation error. (Thomas, Cabrol, Anderson, Grin) Mostly complete, with some continuing analysis. This work has been completed for most of the 2004 study, although there was some delay in returning the rock samples from the field. These samples were again collected in December and January and are currently awaiting spectral analysis at a commercial laboratory. There was also some difficulty in getting the scientists to unanimously commit to a particular interpretation for different environmental features. Consequently, it is unlikely that we will be able to find direct contradictions to scientific findings in the control room. This study is ongoing, however, in parallel with the findings from the second year’s analysis.

1.4 Study how the scientists interacted with the information tools and other artifacts in the control room. Study how the tools meet or did not meet the scientists needs throughout the mission. Determine if training, tools or equipment could have prevented each error and whether the error might be a factor in past or future Mars missions. Create recommendations regarding the information tools provided to the science team and determine what features or functions would be useful in future missions. (Thomas, Coppin). Mostly complete, with some analysis ongoing. The first principle recommendations from the first year of study involved the analysis of the EventScope interface (see Appendix 1) which determined the substantial improvement made when several refinements were made to the system after the first week of operations. The second recommendation involved eliminating the collection and processing of the stereoscopic images, which did not seem to be used heavily in the science analysis (e.g., Glasgow, Pudenz, Thomas, Cabrol, Coppin and Wettergreen (2005), Observations of a Science Team during an Advanced Planetary Rover Prototype Field Test). This controversial conclusion was contested by some of the scientists, but ultimately the engineers down-graded the priority of the stereo-imaging. Although the performance in the final year seems to have been very successful, it will be difficult to document exactly how much of this success was caused by, or perhaps limited by, the decision to down-grade the stereoscopic system since the ultimate performance is confounded by so many other factors. However, we were very encouraged by the fact that our analysis led to specific changes in the rover system based on observations of science effectiveness rather than the scientists self-reports.

1.5 Refine the science information systems (Thomas, Coppin). Complete. Coppin and his team refined the EventScope and website interfaces throughout the mission. Several of these changes were a direct result of the analysis created as part of this project. Most notable among these are the specific benefit observed with adding templates to the programming interface and the creation of a web interface that allowed the scientists to easily access the highest resolution version of the available data.

2005-2006

2.1 Participate in the 2005 field campaign, two to three weeks of science operations at NASA Ames. Video and audio record scientist observations, as before, following interrupt interview protocol, and track information usage with computer-tracking software. (All) Complete. This year we again participated in the LITA mission operations room. Two members of the Iowa team were again present throughout the mission. Based on the observations of the first year, we decided not to transcribe the entire mission, which was very expensive. Instead we are using digital recordings and hope to combine our detailed notes and the fast access to various points of the conversation to substitute for the complete transcripts. This year we also abandoned the interrupt protocol because of the difficulty some scientists faced when asked to commit to a particular testable hypothesis. Instead we have based our analysis on the science reports produced by the science team each day. These reports have proven to be much richer sources of specific observations and hypotheses than the verbal protocols and also have greater validity than comments pulled for casual conversations. We also changed the protocol to include the consistent observation of scientist activity every five minutes throughout the mission operations.

2.2 Travel with science team to Atacama. Visit science sites explored by rover and audiotape and record science team's observations of the sites for comparison with control room observations. (Thomas, Cabrol, Grin and Anderson) This year we made two trips to the Atacama Desert in Chile. The first trip was made before the Carnegie Mellon Team had left the field. Ingrid Ukstins Peate, a University of Iowa geologist, traveled with Thomas to 4 of the rover test sites. Together Thomas and Ukstins Peate developed a protocol to test the various comments made in the transcript and in the science summaries, took reference images of nearly all the rover stopping locations, and collected rock and soil samples for laboratory analysis. Thomas and Ukstins Peate returned in early January to study the 3 remaining rover sites. This trip was immediately followed by the full science team and two more Iowa student team members arriving in Chile. The whole team traveled to all the rover sites. We provided books to the science team members that documented all their science reports, and the key images collected from each site. We also highlighted those sites that were of greatest scientific interest because of interpretations made and verified or contradicted by our earlier site visits. For each week of operations we took the science team to two rover locations and asked them to reflect on approximately 7 features of the locales drawn from the science summaries. We believe that these records and the video and audio recordings made in the field are the first records of a scientist reflecting on their own observations made through a robot. This will form the basis of our analysis for task 2.3.

2.3 Repeat task-analysis and observation-map analysis to determine if significant changes in the scientist performance could be attributed to improvements in the tools the scientists used. (Thomas) Ongoing.
2.4 Validate transcript analysis to verify domain-specific information. Resolve controversial findings with results of laboratory analysis. Trace any errors revealed in the field exercise back to the original data that provided them. (Thomas, Cabrol, Grin and Anderson)
2.5 Create recommendations regarding the information tools provided to the science team and determine what features or functions would be useful in future missions. (Thomas, Coppin)
2.6 Prepare software for the third-year field test. (Thomas, Coppin)
2.7 Identify and prepare logistics for third-year field test. (Cabrol, Grin and Anderson)
2006-2007
3.1 Travel to third-year test site. Collect a data sequence simulating a rover moving about the remote terrain. (Thomas, Anderson, Cabrol, Grin)
3.2 Install field data on three version of science interface software. Recruit three science teams to participate in three sessions, one with each software package, separated by a period of at least one month. Record audio, video and data interaction during the experience.
3.3 Analyze transcripts to determine what data helped the geologists make different observations. Compare these data sources to data sources that are more traditionally build into a rover. (All)

Challenges and Opportunities:

One of the greatest challenges that emerged from this study was the realization that objective truth is more nebulous than we had expected. The fundamental logic of this proposal was that we would compare the observations of the scientists with the "true" observations in the field and laboratory. The difference between the rover-mediated observations and the field observations would represent the "error" which we would then design to correct. Several factors seemed to work against this approach:

1. The scientists often expressed their ideas as a range of possibilities. Sometimes this range was so large that it precluded very little. In some cases it would be nearly impossible to make a field observation that contradicted the science summary.
2. Sometimes the meaning of a word used by one scientist differed from the meaning used by another scientist. Consequently if a scientist in the control room reported that there was a desert pavement habitat, for example, his or her notion of what constituted a desert pavement habitat might be different from a scientist in the field. Consequently, although the difference of opinion might be recorded as an "error," it was merely a difference of meaning.
3. In the field several scientists could look at the same feature and come to different conclusions. For example, at one site the scientists were asked whether there was a paucity of white rocks. Three scientists said that there was a paucity of white rocks. Another, who happened to be very interested in white rocks, and had been moving from one to the next, examining each in turn, reported that there was no paucity of white rocks. Apparently "truth" depends on one's frame of reference.
4. In some cases, when the observation in the field seemed to contradict an observation made in the control room, the scientists searched for a reason for the difference and, instead of reporting that there was a difference in the observations, sought to explain why the differences were there. It is very difficult to separate rationalization from objective observation in these cases.

We will continue to refine our protocols to address these challenges. For example, this summer Thomas has received a grant with Ukstins Peate and another Iowa geologist, Mark Reagan, to develop a system for objective evaluation of geologist performance. We hope that this new assessment technique will help to reduce the challenges presented above. The general strategy is to create a culture in which very specific, low level observations are made and these low level observations are then used to support higher level interpretations. The science team seems very amenable to this approach.

We have also begun plans for the year 3 investigation. The location of this new test site will not be publicly disclosed in order to ensure that the analysts are not able to use information about the general geographic region to improve their interpretations. We hope to visit this new location in August of 2006. We will use a specially equipped camera to simulate a dataset taken by a rover. We will then use these new data sets to simulate an entire rover mission and determine how well a science team can interpret the data from the new site.

Publications and Presentations resulting from this effort

1. Pudenz, E., Glasgow, J., Thomas, G., Coppin, P., Wettergreen, D., Cabrol, N. Searching for a Quantitative Proxy for Rover Science Effectiveness, Proceedings of the 2006 Conference on Human-Robot Interaction, March 2-4, 2006, Salt Lake City, Utah.
2. Glasgow, Justin, Erin Pudenz, Geb Thomas, Nathalie Cabrol, Peter Coppin and David Wettergreen (2005), Observations of a Science Team during an Advanced Planetary Rover Prototype Field Test, Ro-MAN Conference, August 13-15, Nashville, TN, 2005.
3. Thomas, G., Coppin, P., Cabrol, N., Wettergreen, D., Pudenz, E., Glasgow, J. (2005), Collaborative Virtual Environments for Control of Planetary Exploration Rovers, Special Session on Human Robot Interaction, Human Computer Interaction / Virtual Reality Conference 2005, July 22-27, 2005, Las Vegas, NV.
4. G. Thomas, Engineering Robotic Geology for Mars Exploration, IIE Annual Conference, March 2004, Houston, TX.
5. Keynote speech, “Science with Robots on Earth, the Moon and Mars,” ASME Student Leadership Training Seminar, Iowa City, IA 9/24/05.

Appendix A

Template and Positioning Features Have a Notable Impact on Rover Planning Software Package[1]

Erin Pudenz, Justin Glasgow, Geb Thomas
Mechanical and Industrial Engineering
The University of Iowa
Iowa City, IA, 52242, USA
geb-thomas@uiowa.edu

Peter Coppin
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15213, USA
coppin@cmu.edu

Nathalie Cabrol
NASA Ames Research Center/ SETI Institute
Moffett Field, CA, 94035-1000, USA
ncabrol@mail.arc.nasa.gov

David Wettergreen
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15213, USA
dsw@cs.cmu.edu

***Abstract* – This study measures the efficiency of rover task planning in EventScope, a software tool used during a simulated planetary exploration mission. Specifically, it considers how allowing users to indicate positions with virtual pins on a map and providing reusable, modifiable programming sequences, or templates, resulted in an almost four-fold efficiency improvement during the Life in the Atacama rover field experiment. In this experiment a rover controlled by scientists in Pittsburgh, PA spent two weeks searching for potential habitats in the arid Atacama Desert in Chile. The science team directed the rover to autonomously navigate distances of several kilometers between target sites. Task programming efficiency was measured as the time spent using EventScope to create the daily rover program sequence divided by the number of individual rover tasks in the sequence. There was no significant difference in time spent using EventScope between the two conditions, but the science team created more complex programs and were able to enter the programs in quicker when the programming aids were available.**

***Index Terms* – Human Robot Interaction, Mobile Robots, Supervisory Control, Vehicle Teleoperation Interface, Remote Rover Exploration**

I. Introduction

The value of using remote robotics to explore potentially dangerous and distant environments has become more apparent with the recent success of the 2004 Mars exploration mission. Now the focus is on creating more efficient ways to perform these remote missions. Mission success depends on effective communication between the science team and the rover [1, 2, 3]. Much of this communication emphasizes the rover's motion and navigational tasks. Thus, creating an efficient way in which to generate these tasks is essential to a successful mission [4]. Software packages such as EventScope can facilitate this communicate by consolidate the information provided by the rover and put this information in the context of the remote terrain in which the rover is operating. Whereas previous reports of software for rover programming often emphasize the capabilities provided by a particular package [5, 6], this paper considers how changing aspects of a single tool can radically affect tool performance. It also introduces a general metric of performance that will allow other researchers to compare disparate task-level rover programming software. The two versions of EventScope were compared during the two conditions of year two of the Life in the Atacama field campaign. The mission objective was to search for possible life in the Atacama Desert.

[1] This work is partially supported by NASA's Applied Information Systems Research Program under NASA grant NAGW5-11981 and NNG05GA51G and the Astrobiology Science and Technology for Exploring Planets under grant NAG5-12890.

EventScope aided the science team in developing and sending rover plans to the semi-autonomous rover Zoë in the Atacama.

Remote robotics has advanced considerably in space exploration since ROTEX, the first robot that was remotely controlled in space [7]. Rover autonomy is one area that has advanced substantially. Although, intuitively it seems that an increase in autonomy would result in a decrease in human interaction, research has shown that as the autonomy of the rover increases, there is an increased need for human supervision [3, 8]. However, there has yet to be a general way to compare the different software packages that allow for human-rover communication.

In the Atacama mission, scientists in Pittsburgh, PA supported by scientists at other regions in the United States supervised the activities of Zoë. This mission used a supervisory control method which allowed the science team to communicate with Zoë. Supervisory control consists of an operator or "supervisor" who sends a task or sequence of tasks to the rover [9], which autonomously completes the tasks [4].

Here we define a task as an activity or group of closely related activities that describe a single unit of work as envisioned by the rover designers and programmers. The complexity of a task depends on the rover, its design, and its function. During this field test examples of tasks included, traversing several kilometers between locations, collecting images using available filters or camera systems. However, the much less complex task of acquiring a single image from the "workspace camera," mounted below the rover, also constituted a single task. Often a particular task includes multiple specifications of various parameters that define how the task should be executed, such as the determining of the range over which adjacent images should be collected. The complexity of task and the number of parameters it requires may vary from task to task and system to system, although each is clearly defined by the rover and its programming language.

Following the general operations model used by the MER science mission [1], tasks were created and entered into EventScope at the end of the sol (a simulated mission day, 24 hours) by the science team. These tasks were uploaded to Zoë, which would complete the tasks the next sol. The supervisory control approach is well-matched for missions that are restricted by large delays in communication with the rover and missions with small allotments for bandwidth [4]. Both constraints were present during the Atacama mission.

This study compared two versions of EventScope: A and B. Version A was used during week one of the Life in the Atacama mission in September of 2004; Version B during week two of the mission in October 2004. There are two main differences between the versions. Version B includes a positioning feature, or pinning operation, and a template feature that were not in Version A.

In Version A, scientists indicated positions by manually determining and entering in coordinates into EventScope. The pinning operation in Version B, as seen in Fig. 1, allows the scientists to place virtual pins on a satellite image to designate a position. Ref. [1, 3] suggests that allowing individuals to click directly on a map increases the position precision.

The template feature enables the team to save tasks in the interface and use them later to create new tasks. It was observed during the first week of observations that the science team tended to repeat certain sequences of operations, such as collecting adjacent images to form a panoramic view of a site, then collecting a series of images under the rover. Completely specifying these repetitive tasks was a time-consuming and tedious chore, sometimes lasting until 4 in the morning. Version B allowed the scientists to copy and modify task sequences. These new functions are believed to have made EventScope more efficient.

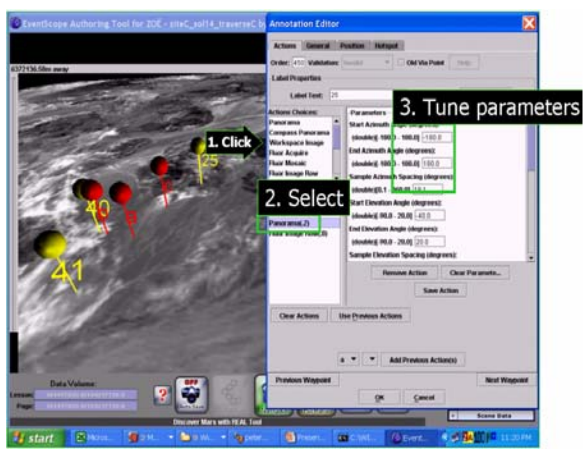


Fig.1 Screenshot of EventScope pinning and template features

This study measures whether the refined version of EventScope (Version B) increased the efficiency of the science team's ability to plan and send rover tasks. The experimental hypothesis is that because of the improvements in the software interface, the science team spent less time programming the rover's daily task sequence with Version B of EventScope than with Version A. Further, the advances in Version B would allow the science team to create more tasks per rover sequence (more sophisticated sequences) than they would with Version A. Finally, given the previous two conditions, it is conjectured that Version B of EventScope will have a smaller time-to-task ratio than Version A.

I. Method

A. Background

During the Life in the Atacama mission of 2004, the rover Zoë explored two distinct arid regions in the Chilean Atacama Desert during two separate conditions. Each region was explored for seven days. The goal of the mission was to look for signs of life in the desert.

The science team was made up of six scientists the first week and eight scientists during the second. Each team was composed of scientists from different disciplines. The first team consisted of three scientists specialized in biology, two in geology, and one in spectroscopy. The second team consisted of three scientists specialized in biology, four in geology, and one in spectroscopy. The actual number of scientists available for any sol varied, but was typically five during the first week and seven during the second week. The team, with the exception of one biologist during week two, worked from a mission control room set up in Pittsburgh, PA where they analyzed the data collected by Zoë and created sol's rover plans. The biologist who was not present communicated with the other scientists via emails and conference calls.

The team received data of the landing site similar to that available during the MER mission [1]. Then, the scientists had until the beginning of the next sol to create and upload the rover task sequence for the following sol's operations. At the end of the following sol, the team would receive the data collected via a password-protected website.

Typically, the scientists received the previous sol's data around 7:00 p.m. and then meet to review the data and to discuss a rough plan for the next sol's operations. After this, the team would break into smaller groups or work individually to analyze the new data. Around 10:00 p.m. the scientists would meet to finalize the rover plan and to enter it into EventScope. The scientists normally finished between 1:00 a.m. to 3:00 a.m. and rested until the 11:00 a.m. the next morning. From 11:00 a.m. to 3:00 p.m., the scientists re-analyzed the data. At around 3:00 p.m., the scientists would summarize their analysis for the group. The team would then break until the new sol data arrived.

The scientists used EventScope to create and upload the rover plan for each sol. After determining the location of interest, the team would specify what data they wanted Zoë to collect, along with any subtasks required to perform the task. These subtasks included, but were not limited to, determining location, camera position, and filter.

Version B of EventScope incorporated two features that were not in Version A. These functions are the template feature and the pinning operation. The template feature allowed the team to store past task sequences. This enabled the team to use stored common tasks to create new tasks without having to retype information.

The pinning operation allowed the scientists to put a virtual pin on a digital elevation map textured with satellite images to designate where Zoë should go to collect data. This eliminated the need to manually determine the coordinates in order to create the task. In Version A, the spectroscopist would determine the coordinates of where the team wanted to perform a task by locating the point on an IKONOS (satellite) image. The IKONOS image had coordinates imbedded into it. However, these coordinates were not in the correct format for the rover plan. A member of the EventScope support team would convert the IKONOS coordinates into the correct format. These converted coordinates were then given back to the science team to be entered into the EventScope rover plan.

During both conditions of the field test, audio, and visual records of the scientists' actions were collected. The experimenters also kept a log recording the scientist's location, activities, and their collaborating groups. This log was updated every five minutes.

At the end of the mission, the website log, the EventScope log, and the rover plans were collected. The website log recorded all data uploads and downloads to and from the password protected mission website. The EventScope log recorded the activities performed in EventScope. The rover plans are the tasks that were given to Zoë to perform requested operations.

B. EventScope Log

The EventScope log was analyzed to determine the amount of time EventScope was used. This log recorded the date, time, and source of each operation along with the tool used and the operation itself. The log was analyzed by dividing it into activity clusters, blocks of time spent working in EventScope. Each cluster was defined as a period of activity with no breaks between actions lasting longer than 5 minutes. The total amount of time on EventScope for each sol was determined by summing the duration of all activity clusters.

C. Rover Plan

The Rover plans were analyzed by counting the number of tasks requested for each sol. Written in XML, these codes included the initial state of Zoë, the desired tasks to be performed, and the final location of Zoë. Each task includes all subtasks that are needed to obtain the operation. Thus, the amount of code required to perform each task was not considered in the analysis.

Due to a clerical error, two versions of the rover plan were available for Sols 3 and 4. Sol 3 had a one task difference between the two versions. The version with the most tasks was used in any calculations regarding Sol 3. Sol 4 had the same number of tasks in both versions. Thus, this number of tasks was used in all calculations.

D. Time-to-task Ratio

The time-to-task ratio was defined to be the duration of the activity clusters divided by the number of rover tasks for each sol. No time-to-task ratios are available for Sols 7 or 14, because these were the last sols of each mission week and no new command sequences were generated.

III. Results

A. EventScope Log

A paired t-test was performed to determine the significance of the time spent using EventScope during each week. The scientists spent the same amount of time programming task sequences with each version of EventScope as seen in Table I, ($t(5) = 0.776$, $p = 0.4731$).

B. Rover Plan

As Table I indicates, the scientists generated sequences containing more tasks when using Version B of the software than when using Version A of the software (paired-$t(5) = -3.85$, $p = 0.01$).

TABLE I
Data Analysis For Each Sol

Sol Number	Total Time on EventScope	Number of Commands in Rover Plan	Time to Command Ratio
1	2:08:19	25.00	0:05:08
2	3:35:18	22.00	0:09:47
3	3:03:56	26.00	0:07:04
4	3:29:42	30.00	0:06:59
5	4:08:46	19.00	0:13:06
6	2:24:00	21.00	0:06:51
7	1:35:15[a]	0.00	N/A
Summation	18:50:01	143.00	N/A
8	3:51:11	104.00	0:02:13
9	1:19:58	129.00	0:00:37
10	2:29:53	64.00	0:02:21
11	4:25:44	94.00	0:02:50
12	2:10:59	142.00	0:00:55
13	1:33:27	25.00	0:03:44
14	0:08:45[a]	0.00	N/A
Summation	15:51:12	558.00	N/A

[a] Not used in calculations because no rover plan was created.

C. Time-to-task Ratio

As both Table I and Fig. 2 suggest, the time-per-task ratio was significantly higher for Version A than Version B (paired-$t(5) = 3.93$, $p = 0.01$).

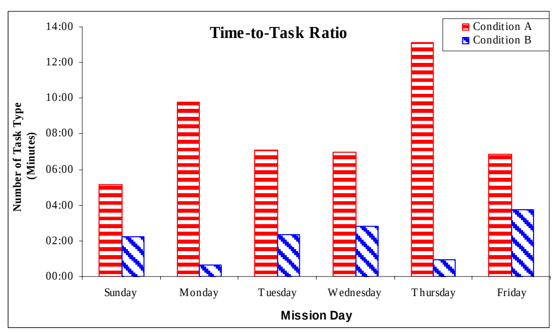


Fig. 2 The Time-to-task Ratio for each mission day with Version A and Version B. Time-to-task Ratio was calculated by taking the time spent in EventScope divided by the number of tasks in the rover plan that was entered into EventScope.

IV. Discussion

A. Results Implications Regarding EventScope

The results indicate that Version B of EventScope was more efficient that the Version A during this mission. The hypothesis that the number of tasks created in the rover plan would be significantly higher in the first week than in second week of the mission was confirmed. Additionally it was found that the time-to-task ratio was also substantially lower for the second week.

Originally it was suspected that the implementation of the positioning pinning operation and the template feature to the EventScope software would increase the speed at which the team could enter the tasks into EventScope and thus reduce the time spent using the software package. However, although the scientists were able to program the tasks faster, the improved efficiency did not reduce the amount of time that the team spent using the software. This disproved the first hypothesis that the scientists would spend less time using EventScope in Version B than in Version A. This deviation from what was originally expected may be explained in several ways. One way is by the team's set schedule. Instead of altering the schedule in response to the efficiency increase, the team maintained the original schedule. As the scientists adjusted to the number of tasks they could enter into EventScope in their allotted time the daily rover tasks became more complex.

Sequences created with Version B contained nearly four times as many tasks as Version A. This substantial increase in number of tasks may have allowed the scientists to have more in-depth data to analyze when determining the presence of life in the Atacama Desert, their primary objective.

The team stayed well into the morning to complete the rover plan and enter it into EventScope. Although the hours were long and tedious, the scientists preferred to stay and complete complex rover plans instead of simple and short plans. This implies that the team felt that the data generated from the more complex plans was beneficial enough to justify the hours spent entering plans.

An alternative explanation for why the time spent in EventScope did not decrease, is that the scientists had more tasks than what they could convey to Zoë in the allotted time. The increase of efficiency of EventScope coupled with the lack of decrease in time using EventScope is consistent with what would be found if the software package was limiting the communication between the scientists and Zoë. However, further research would need to be completed to thoroughly analysis this hypothesis.

The results suggest that Version B is more efficient than Version A. There are several variables in this study that may have contributed to the increase in efficiency besides the added features in Version B. However, these variables would not have resulted in the dramatic increase in efficiency on entering in the rover plan into EventScope.

Version B was used while the science team explored a different area of the Atacama Desert than they had with Version A. In both cases the scientists were trying to identify habitats conducive to life. There may have been some areas that were more interesting in this respect when Version B was used than when Version A was used. These sites may have required more complex analysis and tasks. However, there is no evidence to support this. The transcripts do not suggest that the scientists found one site or the other more interesting nor more demanding of their attention. Certainly, there was no indication that features in the area required four times as many tasks to analyze.

The number of scientists participating varied between the two weeks. During the first week there were six scientists; during the second week there were eight scientists. Some of the scientists participated in both missions while others only participated in one. The increased number of participants in the second week could account for the increased complexity of the rover plan. However, this seems unlikely to explain the large increase in efficiency of entering in the plan into EventScope. Although all of the scientists discussed the plan while looking at the EventScope interface and learned the details of the tasks directly or indirectly from the interface, only one of the participants actually entered in the rover plan each sol. The same participant performed this job with both versions of the software. Thus, the larger number of participating scientists did not affect the number of people working directly with EventScope. A complex plan should have taken the EventScope programmer/scientist longer to enter than a simple plan.

A learning curve might account for the participant's improved efficiency in entering in the rover task sequence, but that seems unlikely. Version B was used after Version A, so it is possible that the participant training rather than software design could account for the improved performance. If this was true, it would be expected that task programming efficiency would increase steadily from the first sol of the first week to the last sol of the second week. There should also be a very small difference in performance between the last day using Version

A and the first day using Version B, if not a slightly lower efficiency when starting with Version B because a change in the software would diminish the benefit of training experience and the performance loss suffered when the participant was not working with the software. The exact opposite pattern was observed. There was little increase in efficiency during each week of operation and there was a substantial jump in efficiency between the two weeks, corresponding to the introduction of the new version. This indicates that the increase in efficiency is not due to a learning curve of the participant entering in the plan.

With the above variables determined not to be the reasons for the increased efficiency, it appears that the efficiency increase is due to the addition of the positioning pinning operation and the template feature implemented in Version B. These functions helped the team with their interaction with Zoë. Although there was no statistical difference between Version A and Version B regarding the time the scientists used EventScope, there was significant difference in the number of tasks in the rover plans and in the time-to-task ratio between Versions A and B.

There are several reasons why the pinning operation may have been effective for the team when entering the rover plan into EventScope. The pinning operation eliminated the need to manually determine and convert the coordinates in the correct form. This operation took the job of determining the coordinates and translated it into determining areas of interest on a map. This is a task that is much more natural for the scientists [1, 3]. The less time that the team has to spend translating information into a format that the rover can understand, the more efficient the system becomes and the more beneficial it is for the operators involved.

The template feature also benefited the team in entering in the rover plan in a more efficient way. It enabled the team to create frequently used subsets of a particular task. This was useful, because the team did not have to decide and reenter the more trivial aspects of a task. The templates allowed the team to organize their rover planning in a pattern that matched the way they thought. They were able to relate it to previous rover plans, which helped with visualization of the requested data and elimination of redundant typing.

B. Results Implications Regarding Methodology

Besides analysis of the efficiency of this particular software package, this study also creates a methodology in which to study the different software packages that are being used in remote rover missions. Specifically this methodology could be used to compare different software packages and versions used to support supervisory control systems in regards to rover planning tasks. Many of the software packages that are currently being used are considerably different from one another [e.g. 1, 3]. These packages use different task codes to communicate the tasks to the rover. These differences create problems in comparing one software package to another. Although there are many differences, every software package communicates information from the operator to the rover. Because the current methodology investigates the general tasks that are given and not the specific format of the tasks, other software packages can be analyzed in the same way. This is an essential tool when determining which software package is best suited to be used in an upcoming mission.

This tool for studying different supervisory control system software packages is important because it can compare the common denominator. However, there are limitations to the current methodology. It is important to look at the complexity of each task. The current method does not take into consideration the number of parameters needs to perform each task. For the current study, this is not an issue because the number of parameters needed to perform tasks does not differ between the different EventScope versions. The current methodology sets a stepping stone to create an even stronger methodology in which to compare supervisory control systems. For future studies, it would be beneficial for the methodology to also include the complexity of completing a task based off of the rover command sequences.

V. Conclusion

A. Rover Planning Software Packages Implications

Rover planning software must be efficient to facilitate communication between scientists and the rover. These software packages play a crucial role in human-rover communication and can greatly influence the success of a mission to explore potentially dangerous or distant environments. The current study found that the efficiency of EventScope has increased by nearly four times since implementing the position pinning operation and the template feature. These features are important not only to EventScope but also to future advancements in rover exploration technology. Implementation of these features into other software programs should improve their efficiency as well. This results in improved software packages that allow scientists to express their requests to the rover, which will product valuable data and more successful missions.

It is also be important to be able to evaluate these different software packages to determine which package is best for a given mission. The methodology used in this paper is a strong starting point to create a way in which to compare rover planning software packages between different versions and packages. However, future research should be completed to improve this methodology.

B. Future of EventScope

With the success of the positioning pinning operation and template function, more features to increase the efficiency of EventScope are being implemented. The current pinning operation is being investigated to see if it is beneficial to improve the access of data by clicking on the pins. There also are plans to make EventScope more useable by adjusting the interface of EventScope to eliminate overlapping text on the pins, as well as, to changing the actions to be listed in chronological order. Other changes include allowing the scientists to view data as it downloads instead of waiting for the data to be completely downloaded.

Entering in the rover plans is not the only way to create a more efficient software package for the. There are current plans to create an area in EventScope that allows the team to know why certain operations were not performed. This will be an important implementation. Ref. [3, 4] suggests that there are significant benefits in allowing the rover to explain what problems have arisen and why the task sequence was not completed. This information will allow the team to make more informed decisions when creating rover plans.

Understanding current and possible future locations is also essential for communication with the rover. Presently, EventScope has a triangulation tool that assists in determining the location of the rover. The tool is being revamped to make it easier to use. This will be very valuable for the operation since the rover's internal localization is primarily an estimate. The next step is to test the tool with additional datasets and verify the results.

With the already implemented improvements and the current plans to create new useful tools, EventScope will be a software package that allows operators to efficiently communicate with remote rovers. This heightened efficiency will create a high grade of performance in communication between the operator and the remote rover.

References

[1] P. Backes, et al. "The Science Activity Planner for the Mars Exploration Rover Mission: FIDO Field Test Results," *Proceeding, 2003 IEEE Aerospace Conference, Big Sky, MT.* 2003, 1-15.

[2] T. Fong, et al. "Operator Interfaces and Network-Based Participation for Dante II," *Proc. SAE Int. Conf. on Environmental Systems*, 1995.

[3] T. Fong, C. Thorpe, and C. Baur. "Advanced Interfaces for Vehicle Teleoperation: Collaborative Control, Sensor Fusion Displays, and Remote Driving Tools," *Autonomous Robots*, 11, 2001, p 77-85.

[4] T. Fong and C. Thorpe. "Vehicle Teleopation Interfaces," *Autonomous Robots*,2001, 11, 9-18.

[5] G. Thomas, A Bettis, N. Cabrol, A. Rathe, and T. Foster, "Analysis of Science Team Activities During the 1999 Marsokhod Rover Field Experiment: Implications for Automated Planetary Surface Exploration," J. Geophys. Res, 106, E4, 7775-7783 2001.

[6] C. Stoker, and B. Hine, "Telepresence control of mobile robots: Kilauea Marsokhod Experiment," presented at American Institute of Aeronautics and Astronautics, Reno, NV, 1996.

[7] G. Hirzinger, B. Brunner, J. Dietrich, and J. Heindl. "ROTEX-The First Remotely Controlled Robot in Space," *IEEE Conference on Robotics and Automation, San Diego*, 1994.

[8] D. Woods, J. Tittle, M. Feil, and A. Roesler. Envisioning Human-Robot Coordination in Future Operations, *IEEE Transactions on Systems, Man, and Cybernetics C.*, 34, 2004, p 138-153.

[9] T. Sheridan. *Human and Automation: System Design and Research Issues.* A John Wiley & Sons, Inc. 2002.

[10] R. Arvidson, et al. FIDO Prototype Mars Rover Field Trials, Black Rock Summit, Nevada as Test of the Ability of Robotic Mobility Systems to Conduct Field Science *Journal of Geophysical Research,* 107 E11,2002, p 2(1-16).